

Testimonial for AI Evals Course

3 MIN · YOUTUBE · [HTTPS://WWW.YOUTUBE.COM/WATCH?V=VKWJPRQFV0K](https://www.youtube.com/watch?v=vKWJprqFv0k)
<https://www.youtube.com/watch?v=vKWJprqFv0k>

SUMMARY

The speaker emphasizes the importance of improving AI agents to enhance user engagement and reduce issues like hallucinations. They recommend a course by Hamill and Treya that provides structured methodologies and practical techniques for building better AI systems, highlighting the significant benefits gained from systematic error analysis and architectural design.

- Importance of rapid improvement in AI agents for user retention and conversions.*
- The three goals framework: comprehension, specification, and generalization for addressing challenges.*
- Systematic error analysis can lead to significant reductions in hallucination rates.*
- Use of axial coding to categorize failure modes for better understanding.*
- Importance of statistical methodologies in evaluation processes.*
- Techniques for bias correction to enhance confidence in automated evaluations.*
- Architectural design insights for cost optimization, potentially reducing inference costs by up to 85% while maintaining accuracy.*
- The course focuses on practical techniques and mental models rather than just tools, providing a comprehensive framework for improvement.*

Do you want to be in the top 5% of AI teams? Lots of teams are struggling with how to rapidly improve their AI agents for reduced hallucinations, improved recommendations, user engagement, retention, ultimately conversions. I think about it like this. If you went to a coach or therapist and within the first minute there was a bad interaction, would you pay them? No. And your users wouldn't either in terms of time or money with your agent. Everyone's talking about how evolves are the moat in AI, but few are giving strategic, technically sound, and tactical advice and processes like Hamill and Treya are in their AI evals course. I was super fortunate to take the course, learned a ton, and I highly recommend every team building agents to do the same. Your team, your product will become better, and your customers will be happier for sure. So, here's some of the stuff that I learned. If you end up taking the course, highly recommend it. This is what you could expect at a high level. I really love learning about the three goals framework. So, comprehension, specification, and generalization that gave us a structured way to understand our challenges and address them methodically. It's also a good mental model. What really resonated was the systematic error analysis process that was super scientific and measurable.

They gave a systematic approach to do open coding error analysis and that showed me that this was perhaps singularly the most beneficial step function improvement to your agent hallucination rates can drop by double digits just from this alone is something I learned from the course. So we learned proper methods to identify patterns to then use axial coding to categorize failure modes into a taxonomy. How do you get to the you real user queries fast and improve the agent fast?

The lessons on LLM judges applied to a few specific scenarios. We were able to focus on building judges that targeted specific failure modes rather than attempting to evaluate everything all at once. So I learned that following a statistical methodology with the proper train dev test splits for the data sets for the evals. Super important. Learned how to implement bias correction and that will significantly improve our confidence in automated evals. So I'm pretty stoked about that too. And then rag. So some thought leadership to really get what are the experts saying about this is Rag actually did. And we even talked about architecture design which was super cool and it gets into the nitty-gritty of how you can design an LLM application architecture for cost optimization. So we went into nitty-gritty stuff like how might we reduce inference costs by up to 85% while maintaining 95% of the previous accuracy. That alone made the course worth every penny. Right? If you could do that and your costs go down but your accuracy is the same. Amazing.

That's all to say those are some of the snippets of what I learned from this course. Loved taking it. I learned a lot and I feel like a wizard now to be honest in building agents and LM applications. So if you're struggling with building agents, LM applications, you're struggling with consistent quality, you're not sure how to systematically go about improvement, this course provides a systematic framework and the practical techniques you need to make your agent better. It's not all about the fancy tools, though.

The experts, Hamill and Shrea, also have some of the best insider knowledge I've seen on what the fancy tools out there are. It's really about the process, and it's really about the thinking and the mental models, and that's exactly what this course delivers on. My only regret is not taking it sooner. So, don't be me. Take it sooner.